1-1-1976

Examination of the status of probability sampling in the courts

Boyd Randall

Paul Frishkoff

6

An Examination of the Status of Probability Sampling in the Courts

Boyd Randall
Brigham Young University

Paul Frishkoff
University of Oregon

Auditing researchers, staggering under an ever-increasing blizzard of pronouncements, articles, and memoranda, might well be disposed to believe that complexity feeds on itself. For example, two themes which were of much concern—perhaps even dominant—at the first two University of Kansas Auditing Symposia[1] were the increase in litigation against auditors and the growing sophistication of statistical sampling techniques in auditing. This paper attempts to examine the interaction between the two.

Indeed, his paper arose directly as a result of the 1974 Symposium. At that conference, one of the present authors raised the question of whether statistical sampling (i.e., probability sampling) would be a better defense in the courts than judgment sampling. There appeared to be some division among practitioners present at the conference.

Practitioner Opinions

Some felt that the use of probability sampling would ameliorate the position of the defense in a lawsuit, since probability sampling is viewed as more "scientific," as encompassing more up-to-date technology, and as more susceptible to mathematical "proof" of its validity.

Others were more skeptical or at least agnostic. They felt that expert witnesses might debate the merits of *a particular* probability sampling plan to the ultimate utter confusion of jurors and jurists. Their contention was that the "expert judgment" of a highly trained professional, on the other hand, was less suspect and less susceptible to point-by-point rebuttal.

Our purpose in this paper is twofold. First, we will examine past court decisions to discover whether the courts have shown any preference for probability sampling. Then we will examine cases in which probability sampling was used, either by one of the parties or by the courts themselves, to see what implicit standards for such sampling may be emerging from the judicial process.

For readers who are curious about the outcome of these questions, yet less than enthusiastic about wading through the details of this paper, we will admit at the outset that our conclusions are more equivocal and more tentative than

either they or we might have wished. Auditing remains an inappropriate field for uncertainty-avoiders!

Some Definitions

Probability sampling plans, to quote one source implicitly used by the courts, ". . . make use of the theory of probability to combine a suitable procedure for selecting sample items with an appropriate procedure for summarizing the test results so the inferences may be drawn and risks calculated from the test results by the theory of probability. For any given set of conditions there will usually be several possible plans, all valid, but differing in speed, simplicity, and cost."[2] Simple random sampling is but one of many possible applications of probability sampling.

Judgment (non-random) sampling plans, on the other hand, ". . . have one common characteristic: the probability that an individual (item) is included in the sample is unknown. . . . When the determination of the individuals (items) to be included in a sample involves personal judgment, one cannot have an objective measure of the reliability of the sample results, because the various individuals (items) may have differing and unknown chances of being drawn."[3]

A *census* involves the examination of all items in a population, rather than a sample thereof. The economics and timing of such procedures preclude their use for more facets of an audit.

Extent of Use of Probability Sampling

The official policy of the AICPA has been to condone both probability and judgment sampling.[4] An examination that we made of a (non-random) sample of six recent auditing textbooks indicated that judgment sampling was not explicitly condemned, but the preponderance of material under the heading "sampling" dealt with probability sampling.

Most accredited Schools of Business have required one or more courses in statistical theory and techniques for at least the past decade or two. All national CPA firms and the AICPA have offered training programs or modules on probability sampling as well. One might expect, then, that probability sampling is used overwhelming. This is apparently not the case, however.

A questionnaire survey by Jacobs[5] of CPA firms in the Los Angeles area revealed a wide disparity in the use of statistical sampling within a given office, among firms of a given size, and among firms of differing sizes. As might have been expected, the use of statistical sampling by national firms was far greater than by locals, but the use was far from universal in any size grouping.

A more extensive questionnaire survey, undertaken by Strawser and Hubbard[6] confirmed these findings; their research also indicated that utilization of statistical sampling techniques has been increasing fairly rapidly in recent years. These results were consistent with an unscientific face-to-face inquiry of various Portland, Oregon practitioners undertaken by one of the present authors during the past year.

Judgment sampling, then, is still very much alive, and its acceptability in the courts is thus not a trivial problem.

Survey of Cases

We surveyed a wide variety of cases where statistical sampling was an explicit issue or in which it was mentioned by the courts. We wish that we could report that we found *The Definitive Audit Case* or a truly helpful case of any kind. Instead, we found that references to sampling usually arose where the technique was being used to gather evidence for the court, rather than where it had been used prior to the litigation, such as in an auditor's test of transactions. Therefore, we concentrated on court pronouncements that suggested standards that *should be used* in sampling. Such standards, we argued, should be applicable in almost any context, and in particular where an auditor's use of a sampling method might be questioned relative to the standards recognized by the courts.

Implicit Court Standards for Sampling in General

There is no codified law regarding standards for sampling. Yet standards of other bodies, while not of themselves court standards, become so indirectly when they are recognized and cited as such by the courts on sufficient occasions.

"Handbook of Recommended Procedures for the Trial of Protracted Cases" is a Report of the Judicial Conference Study Group on Procedure in Protracted Litigation. This report was adopted by the Judicial Conference of the United States in March, 1960. It reached the following conclusions on the question of sampling:

> Scientifically designed samples and polls have received increasing acceptance in recent years in government and in industry. The important question to be considered in a given case is whether the contemplated or proffered sample or poll is admissible under existing rules of evidence.
>
> Samples [confined to observable facts]. When a sample is offered through the testimony of the sampler, the report on the sample examined (i.e., on the count of units or the test borings, in the samples noted above) usually does not involve hearsay. In order to project this report, however, the burden of proof rests upon the offeror to show, by the testimony of a statistical expert, that the sample was selected in accordance with accepted principles of sampling so that it properly represents the universe. *Note.* Once this is established, there remain only questions of relevancy, materiality and weight.[7]

The note in the above quote refers to two sources of accepted principles of sampling. They are:

1. Munitions Board Standards, Agency of the Department of Defense, and
2. The American Society for Testing and Materials.

We were unable to locate the Munitions Board Standards, but we did examine the relevant standards of The American Society for Testing and Materials (ASTM) in some detail.[8]

A perusal of ASTM's standards for statistical sampling reveal—by the sophisticated nature of the language, mathematical notation, and footnote references—that these standards were enacted by statisticians for use by those with considerable grounding in the field of probability and statistics.

ASTM's pronouncements also reveal, time and again, a pronounced preference

for probability sampling and an aversion to wholesale nonrandom sampling. For instance:

> In order to make any estimate for a lot or for a process, on the basis of a sample, it is necessary to select the units in the sample at random. . . . The only universally acceptable definition of a random selection is is one made by the use of random numbers, which in effect is the guarantee of thorough stirring of the sampling units in a lot.[9]
>
> Selection by use of random numbers need not be more onerous or costly than hit-or-miss methods of sample selection, provided the sampling plan is thoughtfully formulated. . . . Randomness is obtained by positive action; a random selection is not merely a haphazard selection, nor one declared to be without bias. . . . However, mechanical selection is still usually preferable to a judgment-selection.[10]
>
> The foregoing paragraphs do not mean that nonrandom and judgment samples, for example, may provide useful information for the efficient design of a probability sampling plan. . . . It also should be noted that judgment plays an important role in the *design* of a probability sampling plan. For example, it may be used to assess costs, to estimate spreads and likely values of variances; also definitions of strata. In the actual probability sample, however, judgment is not used in the selection of the individual items of the sample, nor in making the inferences, nor in calculating the risks of decisions based wholly on the sample or succession of samples.[11]

Thus, nonrandom sampling is relegated by ASTM to a role of amassing initial evidence when little is known about the underlying population. Such might be the case in a preliminary stage of an audit of a previously unaudited client, for instance, but would not likely be the case in the *actual selection* of receivables to be confirmed for the engagement, and especially would not be supportable for an engagement that has been of long standing.

The courts have cited the Handbook, and thus implicitly the ASTM standards, on at least three occasions. None involved auditing, however, and none involved a head-on-head confrontation between probability and judgment sampling.

In *Bank of Utah v. Commercial Security Bank*,[12] the courts rejected plaintiff's sampling plan in an action under the Sherman Act for treble damages and injunctive relief. The court held that the restraint involved in a "no check" plan whereby school employees were required to take some affirmative action to transfer funds from defendant bank to a bank of the employee's choice was not shown to be unreasonable.

> . . . (T)he tendency is to admit the results of properly conducted surveys for whatever they are worth in spite of the hearsay difficulty. . . . In this case, however, we do not even reach the hearsay question as it relates to the admissibility of surveys, for we think the trial court was well within its bounds of discretion in refusing to admit a poll conducted as was this one.
>
> A survey is inadmissible when the sample is clearly not representative of the universe it is intended to reflect. See Hawley Products C. v. United States Trunk Co., 1 Cir., 259 F. 2d 69, 77; Handbook of Recommended Procedures for the Trial of Protracted Cases, 25 F.R.D. 351,429; . . . Here the universe was either all school board and hospital employees under the

plan, or at a minimum all school board employees under the plan. The sample was chosen from neither of these groups but instead from those employees who at one time banked with appellants and later switched exclusively to Commercial Security.[13]

In another case, *Berman v. New Hampshire Jockey Club, Inc.*,[14] the court rejected a poll as evidence, citing the Handbook, and finding the poll to be entirely inadequate. One reason was because the sample size was less than 1% of a universe of 3,500 to 5,000. It is of interest that the ASTM standards include a section replete with mathematical notation entitled "Equations for Calculating Sample Size," with the strong implicit assumption that probability sampling is used and that characteristics of the underlying probability distribution are known or can be estimated.[15]

In a third case in which the Handbook was cited by the court, *Grotrian, Helfferich, Schuz, etc. v. Steinway and Sons*,[16] the court held that the survey offered in evidence met the criteria for admissibility of samples found in the Handbook.

Court's Examination of Underlying Probability Distribution

The underlying probability distribution on which a sampling plan was based arose in two cases involving depreciation, both, however, concerned tax litigation, not financial auditing.

In *Commissioner v. Indiana Broadcasting Corp.*,[17] the question was whether a television network affiliation for a two-year term, which is automatically renewable in the absence of termination by the affirmative act of either of the parties, for an unlimited number of successive two-year terms, is a depreciable asset. The Seventh Circuit Court held that it was not a depreciable asset.

> The theory of the statistical tables compiled was that an annual rate of contract termination for each pertinent period could be obtained by dividing the total number of years commenced by all the affiliation contracts during a given period into the total number of contract terminations occurring during the same period. Using that termination rate, taxpayer's expert witnesses testified that the average life expectancy of any given contract could be determined by applying the Poisson-Exponential Theory of Failure. The crux of that theory is that the percentage of failure of items to which it is applied is a constant. . . .
>
> Adopting that theory, and applying it with some modification to the statistical history, the Tax Court found that an estimated useful life of the WISH and WANE CBA affiliation contracts could be determined with reasonable accuracy, and that use of a straight-line method over 20 years was a reasonable basis for depreciation of the contracts.
>
> We think the Tax Court erred. . . .[18]
>
> It is likewise clear that the stipulated exhibit graphically refutes the existence of a basic premise upon which the Poisson Theory relies, namely, that the life expectancy of all contracts is the same regardless of the length of duration of the contract.[19]

Super Foods Services, Inc. v. United States[20] is similar to *Indiana Broadcasting*. Super Foods owned franchise contracts with grocers; they attempted to depreciate the contracts, but the IRS claimed that the contracts had indefinite

lives. The plantiff introduced a study of 486 contracts; the study tended to show that any contract which had been in force more than 12 months had an average life of 86 months. Expert testimony was that the contracts "display a definite and consistent pattern of termination." The government relied on *Indiana Broadcasting* for the proposition that statistics of past performance do not reliably predict termination dates. In ruling for the taxpayer, the court distinguished *Indiana Broadcasting* because in the present case the IRS had introduced no evidence of their own to controvert plaintiff's statistical evidence.

Court's Examinaion of Sampling Plan Used by Litigant

Johnson v. White[21] was an action by Connecticut welfare recipients against the state commissioner of welfare. The commissioner had converted the FADC program to a "flat grant" system to simplify AFDC payments by averaging budgeted needs of each size of assistance unit and making identical payments to each family of a given size. The court held that such averaging was permissible, and explained:

> The defendant adopted a pre-sample confidence interval of 95%, plus or minus 10%; and a post-sample check determined that the selected level of confidence had in fact been met. The defendant's sampling technique was approved by HEW. The plaintiff's claim that in certain components of need for certain assistance unit sizes, the sample size was too small to guarantee the selected level of confidence. The defendant's confidence level, however, was for the average of budgeted needs as a whole, rather than for each component in the standard of need. Giving due weight to the expert opinion of HEW, the court finds that the defendant's sampling technique was adequate.[22]

Use of Sampling Techniques by Court Itself

In *Rosado v. Wyman*,[23] also a welfare case, New York welfare recipients claimed that the state impermissibly lowered the standard of need by eliminating items arbitrarily in 1970. Social Security Act Sec. 402 (a) (23) provides that a state may not lower its standard of need by arbitrarily eliminating items which were included before enactment of the amended statue in 1969.

The court used statistical sampling techniques to determine if amounts paid under the 1970 plan accounted for certain special need grants paid in 1968. The amount of aid received by a sample of 1968 recipients was compared to the proposed 1970 schedule. The court found that 1968 recipients received more than was allowed by the 1970 schedule, refuting New York's allegation of paying 100% of the standard of need in both years.

Some Other Illustrative Cases

Many anti-trust cases employ statistics. In bank merger cases, such as *United States v. Philadelphia National Bank*[24] and *United States v. Manufacturers Hanover Trust Co.*,[25] the government used rather simple arithmetical statistics to show the effect of mergers. For a detailed analysis of these two cases, and—in particular—for suggestions for the use of statistics in similar cases, see Lozowick et al.[26]

An older case involving statistics, a landmark anti-trust action, was *United States v. United Shoe Machinery.*[27] The court discussed minor problems with the government's choice of universe and of sample selection, but accepted the sample pragmatically, because ". . . if anti-trust trials are to be kept manageable, samples must be used, and a sample which is in general reasonable should not be rejected in the absence of an offer of a better sample."[28]

In *Hawley Products Co. v. United States Trunk Co.*[29] (a case, incidentally, cited by the court in *Bank of Utah v. Commercial Security Bank*, which we discussed earlier) and in a related case, *American Luggage Works, v. United States Trunk Co.*,[30] plaintiff attempted to prove that its design for suitcases had acquired a secondary meaning. The survey was "inadmissible to show that in the market of ultimate consumers the plaintiff's design had acquired a secondary meaning when the universe surveyed consisted of retailers."[31]

In *International Milling Co. v. Robin Hood Popcorn Co.*,[32] the evidence included a consumer action survey designed to determine whether purchasers associated the product packaged by International Milling with the product packaged by Robin Hood. The Commission went into detail in describing the survey, the questions asked, the selection of a sample and the standard deviation expected.

Readers may be familiar with the use of sampling in cases where a defendant or plaintiff has asserted that a jury is racially imbalanced and thus in violation of Fourteenth Amendment rights. In cases such as *Carter v. Jury Commission*,[33] courts have been willing to accept a statistical analysis of the population and use that probability to determine that a certain proportion of jurors should be of a particular race. (Such cases, of course, involve arguments from the population, rather than from a sample.) A recent article on jury selection goes into detail on the statistical problems involved.[34]

Sampling Standards Employed by Internal Revenue Service

The IRS issued some rather lengthy standards of sampling in conjunction with the timing of trading stamp redemptions. These standards explicitly condone only *probability* sampling. Thus, "the taxpayer may use any sampling procedures that are in accord with generally accepted probability sampling techniques. . . . While no specific requirements are established for the sampling expert responsible for the design of the study, it is recognized probability sampling is a highly specialized field and that otherwise competent statisticians may not be qualified in the field of probability sampling. . . . The sampling plan which is used must conform to the standards of the 'Report of Committee on Standards of Probability Sampling for Legal Evidence—Admissibility of Data from Probability Samples.' "[35]

This report, published by the Society of Business Advisory Professions, Inc., in cooperation with New York University, is remarkably similar in tone and substance to the previously cited ASTM documents. The report does not allow for the possibility of judgment sampling, and indeed makes the solemn assertion (in two different places) that: "The interpretation of a statistical calculation such as the standard error is not a matter of opinion, nor of judgment, but is a mathematical consequence."[36]

Approval of Probability Sampling by U.S. Congress

In a bill approved in 1964, the House of Representatives specifically permitted the use of statistical sampling procedures in the examination of vouchers. A detailed report accompanying the bill discusses the appropriate use of statistical sampling in the examination of vouchers and the reasons supporting increased governmental adoption of statistical analysis.[37]

A Most Tentative Conclusion

No absolutely definitive case has yet arisen to demonstrate that judgment sampling will not be allowed by the courts in auditing. Nonetheless, some trends seem apparent to us. The courts—and such quasi-judicial agencies as IRS—are relying more heavily on sophisticated statistical documents whose standards barely condone, and certainly do not encourage, nonrandom sampling *when probability sampling is feasible.*

We can not conclude with the lame epithet that "further research needs to be done." There is no more "data" left to examine, and additional "research" will be effected by the courts themselves. Our murky crystal ball suggests that a landmark case involving sampling in an auditing context will arise within the next few years. We caution those auditors who continue to use judgment sampling in the presence of feasible probability sampling procedures to be prepared to defend their logic in so doing. For that matter, given some of the decisions cited, we urge auditors who use probability sampling to be prepared to defend their particular sampling plans in terms of demonstrating that their sample results can indeed be argued to be representative of the underlying universe.

Addendum: Sampling Lost, Sampling Vindicated?

The following case has been verbally cited to us by several colleagues. So far as we know, it has not been officially reported. Moreover, it is somewhat dated and was tried on a municipal court level only. Nonetheless, for the sake of completeness, we include it for the interested reader.

In Sears Roebuck and Co. vs. the City of Inglewood,[38] Sears sued for a sales tax refund of $27,000. Sears had found an error in their own local records as to the extent of out-of-city sales, such sales not being subject to the municipal tax. In support of their contention, Sears conducted a random sample of sales slips, which indicated that the refund ought to be $28,250, plus or minus $4,200, at a 95% confidence interval.

The judge ruled against *any* sampling technique! This ruling was not predicated on any statistical oversight in Sears's procedures, but on the judge's contention that the amount of the refund must be determined exactly, without *any* possible error in estimation, however small. That is, he insisted, as required by the applicable statute, that Sears demonstrate a lack of sales tax liability for each invoice!

Since the Court permitted Sears to perform a "complete audit," the plaintiff did not appeal the case. The complete audit, of which the sample had constituted only 4%, revealed a liability of $26,750.22, well within the confidence interval.

As we said, we doubt that this case constitutes an important precedent, since

special conditions were present which the judge interpreted to preclude *any* form of sampling. If sampling *is* applicable, judges can be expected to follow the precedents that have been cited, and even if a judge does not understand all of the ramifications of sampling, sampling is at least likely to be tolerated.

It is of interest, however, that the author who cites this case goes on to comment, hypothetically, ". . . if (the judge) had refused to admit the amount of the claim as $27,000 because it was based on a *judgment sample* as opposed to the random sampling insisted on here, that also would have been a valid objection."[39]

Footnotes

1. Howard Stettler, Editor, *Auditing Looks Ahead,* University of Kansas School of Business, 1972, and *Contemporary Auditing Problems,* University of Kansas School of Business, 1974.
2. American Society for Testing and Materials, "Standard Recommended Practice for Probability Sampling of Materials." [Designation E 105-58 (Reapproved 1970)], Section 2.
3. Morris H. Hansen, William N. Hurwitz, and William G. Madow, *Sample Survey Methods and Theory,* John Wiley and Sons, New York, 1953 (reprinted 1964), v. I, p. 9.
4. Committee on Auditing Procedure, AICPA, *Codification of Auditing Standards and Procedures* (SAS No. 1), 1973, Sec. 320A.05.
5. John Jacobs, "Statistical Sampling—Is It Being Utilized?" California CPA Quarterly, March 1971, pp. 9-10, 12, 14-16, 33.
6. Thomas M. Hubbard and Robert H. Strawser, "The Auditor and Statistical Sampling," *The CPA Journal,* August 1973, pp. 670-673.
7. 25 F. R. D. 351, 1960, pp. 426-7.
8. American Society for Testing and Materials: "Standard Recommended Practice for Choice of Sample Size to Estimate the Average Quality of a Lot or Process" (Designation E 122-72); "Standard Recommended Practice for Dealing with Outlying Observations" (Designation E 178-68); "Standard Recommended Practice for Probability Sampling of Materials" [Designation E 105-58 (Reapproved 1970)]; and "Standard Recommended Practice for Acceptance of Evidence Based on the Results of Probability Sampling (Designation E 141-69).
9. ASTM, "Choice of Sample Size . . .", *op. cit.,* Sec. 9.1.
10. ASTM, "Probability Sampling, . . ." *op. cit.,* Secs. 7.1.7, A1.2, A1.3.
11. *Ibid.,* Secs. A1.7, A1.8.
12. 369 F.2d 19 (10th Cir. 1966).
13. *Ibid.,* p. 27-28.
14. 324 F. Supp 1156 (D.N.H. 1971).
15. ASTM, "Choice of Sample Size . . ." *op. cit.,* Sec. 5.
16. 365 F. Supp. 707 (S.D.N.Y. 1973).
17. 350 F. 2d 580 (7th Cir. 1965).
18. *Ibid.,* p. 582-3.
19. *Ibid.,* p. 585.
20. 416 F. 2d 1236 (7th Cir. 1969).
21. 353 F. Supp. 69 (D. Conn. 1972).
22. *Ibid.,* p. 75.
23. 437 F.2d, 619 (2d. Cir. 1970).
24. 374 U.S. 321 (1963).
25. 240 F. Supp. 867 (S.D.N.Y. 1965).
26. Lozowick, Steiner, and Miller, "Law and Quantitative Multivariate Analysis: An Encounter," 66 Mich. L. Rev. 1641 (1968). Some other good, recent articles linking law and statistics are: Michael O. Finklestein, "The Application of Statistical Decision Theory to the Jury Discrimination Cases," Harvard L. Rev. 338 (1966); Leo Katz, "Presentation of a Confidence Interval Estimate As Evidence in a Legal Proceeding," *The American Statistician,* November 1975, pp. 138-142; and Joseph G. Van Matre and William N. Clark, "The Statistician as Expert Witness," *The American Statistician,* February 1976, pp. 2-5. (Our thanks to Robert Hamilton for calling our attention to these latter three sources.)
27. 110 F. Supp. 295 (D. Mass. 1953).
28. *Ibid.,* p. 305-6.
29. 259 F. 2d 69 (1st Cir. 1958).
30. 158 F. Supp. 50 (D. Mass. 1957).
31. Hawley Products . . . , *Supra* note 23 at 77.
32. 110 U.S.P.Q. 368 (1956).

33. 396 U.S. 320 (1970).
34. Comment, "The Civil Petitioner's Right to Representative Grand Juries and a Statistical Method of Showing Discrimination in Jury Selection Cases Generally," 20 U.C.L.A. L. Rev. 581 (1973).
35. Rev. Proc. 72-36, 1972-2 CB 771, 779.
36. *Ibid.*, p. 798.
37. House Miscellaneous Reports on Public Bills, Report No. 1643 1/4.6 74/7: V94; Report to accompany H.R. 10446.
38. R. Clay Sprowls, "The Admissibility of Sample Data into a Court of Law: A Case History." *UCLA Law Review,* February, 1957, pp. 222-232.
39. *Ibid.*, p. 230.